Title:

The association of proximal elements of social disadvantage with children's language development at 2 years: an analysis of data from the Children in Focus (CiF) sample from the ALSPAC birth cohort

Authors:

James Law[1], Judy Clegg[2] , Robert Rush[3], Sue Roulstone[4], Tim J. Peters[5]

Affiliations:

[1] Institute of Health and Society, School of Education, Communication and Language Sciences, Newcastle University, UK

[2] Department of Human Communication Sciences, University of Sheffield, Sheffield, UK

[3] Queen Margaret University, Edinburgh, UK

[4] Faculty of Health and Applied Sciences, University of the West of England, Bristol, UK

[5] School of Clinical Sciences, University of Bristol, Bristol, UK

Address for correspondence:

Professor James Law, School of Education, Communication and Language Sciences, Newcastle University, Newcastle-upon-Tyne, UKNE1 7RU T: 0191 208 5250

E: james.law@ncl.ac.uk

Acknowledgements

We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

We would like to thank the ALSPAC speech team who collected the speech and language data and to Laura Miller who has acted as our 'data-buddy' throughout this project.

The UK Medical Research Council, the Wellcome Trust and the University of Bristol provide core support for ALSPAC.

This work was supported by the Department for Education in England.

ABSTRACT

Background

An association between social disadvantage and early language development is commonly reported in the literature, but less attention has been paid to the way that different aspects of social disadvantage affect both expressive and receptive language in the first two years of life.

Aim

This study examines the contributions of gender, parental report of early language skills, and proximal social variables (the amount of stimulation in the home, resources available to the child and the attitudes/emotional status of the primary carer and the support available to him/her) controlling for distal social variables (family income and maternal education) to children's expressive and receptive language development at two years in a community ascertained population cohort.

Methods and Procedures

Data from 1,314 children in the Children in Focus (CiF) sample from the Avon Longitudinal Study of Parents and Children (ALSPAC) were analysed. Multivariable regression models identified the contribution of proximal (what parents do with their children) measures of social disadvantage adjusting for more distal (e.g., family income and material wealth) measures as well as early language development at 15 months to the development of verbal comprehension, expressive vocabulary and expressive grammar (word combinations) at 2 years of age.

Outcome and Results

In the final multivariable models gender, earlier language and proximal social factors, co-varying for distal factors predicted 36% of the variance for expressive vocabulary, 22% for

receptive language and 27% for word combinations at two years. Language development at 15 months remained a significant predictor of outcomes at 24 months. Environmental factors were associated with both expressive scales but the picture was rather more mixed for receptive language suggesting that there may be different mechanisms underlying the different processes.

Conclusions and Implications

This study supports the argument that social advantage makes a strong contribution to children's language development in the early years. The results suggest that what parents/carers do with their children is critical even when structural aspects of social disadvantage such as family income and housing have been taken into consideration although this relationship varies for different aspects of language. This has the potential to inform the targeting of public health interventions focusing on early language and preliteracy skills on the one hand and home learning environments on the other and, potentially, the two in combination.

Key words: expressive language, language comprehension, parents, social disadvantage, Avon Longitudinal Study of Parents and Children (ALSPAC).

What we know about this subject

The level and rate of language development in early childhood is commonly associated with the level of social disadvantage. However, it is less clear whether it is what parents do with their children in terms of optimising the child's communicative environment or whether it is just structural disadvantage that makes the difference.

What this paper adds:

This study shows that there are both shared and different elements in the children's environment which are associated with different aspects of early language development. The results suggest that an intervention focus on critical aspects of the home learning environment could have a considerable effect on the language skills of two year olds which, in turn, has the potential to lead to positive downstream effects on school readiness. This has implications for the model of public health interventions provided by health visitors and speech and language therapists and suggests that it may be possible to quantify the amount of change that is achievable from such interventions.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

INTRODUCTION

Social disadvantage is known to impact on children's language development in the preschool years (Hoff 2006; Huttenlocher *et al.* 2010; Law *et al.* 2011; Roy and Chiat 2013*)* and in later adolescence (Spencer *et al.* 2012). Hart and Risley's seminal work *Meaningful Differences in the Everyday Experiences of young American Children* (Hart and Risley 1995) has been instrumental in driving an understanding in the US about the widening gap in the language skills of very young children in white collar, blue collar and families in receipt of welfare. This, in turn, has led to a focus from the academic and philanthropic sector, for example the Big Word Gap campaign http://www.bwgresnet.res.ku.edu/ and the Thirty Million Word Initiative (http://thirtymillionwords.org/) and the need highlighted by the Clinton Foundation (Clinton Foundation 2013) (https://www.clintonfoundation.org/blog/2013/10/03/closing-word-gap). Although the concerns are undoubtedly real there is a danger of catastrophizing the issue based on a study which only included 42 children, albeit 42 children whose language had been very closely observed for the first three years of their lives. Hart and Risley were clear that the cause of the differences was the verbal input that the children received. However, there are many other issues about the relationship between parental input and material deprivation which we need to better understand if we are to move from a recognition of a concern to evidence based recommendations as to what should be done about it.. For example, one of the key questions for researchers, and indeed practitioners, is how much variance is explained by parental input relative to broader social risks such as income poverty, social class or social deprivation. In an important relatively recent study Huttenlocher *et al.* (2010) suggested that the nature of caregiver speech to children partially mediated the relationship between social background and the children's vocabulary development.

The method for measuring social disadvantage is important here. A distinction has long been drawn between proximal and distal measures of the child's environment (Jessor and Jessor 1973; Schoon 2006) although the distinction in terms of causation has not gone unchallenged (Krieger 2008). Researchers often use distal measures of social status such as family income and maternal education rather than trying to differentiate between what parents/caregivers have (family income and maternal education) from what they do with their child/children in terms of creating communicative supportive environments that foster a child's language development. Such aspects include the presence and use of books in the home (Karras and Braungart-Riekar 2005), how much the child is taken to the library, the material resources available, maternal attitudes to learning and whether parents/carers consider it appropriate to "teach" the child. They also commonly include the amount of TV watched in the home (Close 2004), although the contribution of TV watching and more recently engagement with screens is unclear (Foster and Watkins 2010). Similarly, maternal social support, enjoyment of parenting, and maternal mental health are also implicated (Paulson *et al.* 2009).

In recent years, there has been a move away from studying highly selected or "clinical" samples to using general population studies to investigate language development over time (Law et *al.* 2015; Whitehouse *et al.* 2011; Rescorla *et al.* 2007). The strengths of such population-based studies are in their size, scale and capacity to understand interactions between different aspects of development over time. Investigating the role of social disadvantage in children's early language development in a population cohort offers an opportunity to understand how this association operates over time in the general population as well as the more clinically defined and smaller samples previously studied.

One of the first studies to address the prediction of language at 24 months at a population level suggested that significant social factors accounted for 4.3% and 7.0% of the variation in the 24-month Communication and Symbolic Behavior Scales (CSBS) and Communicative Development Inventories scores, respectively (Reilly *et al.* 2007). Lower maternal vocabulary and older maternal

age were associated with performance on the CSBS and birth order and non–English-speaking background were also associated with vocabulary development. These authors suggested that identification of children with difficulties based on such risk factors, given the relatively low variance accounted for was unlikely to be a promising direction of travel, a rather different conclusion from that reached by Hart and Risley.

A key issue is the methods used to measure children's early language development. In the first two years of life, expressive vocabulary is key as an indication of a child's development and many studies confine themselves to this single measure but by two years we are also interested in early grammatical development which is often captured as early word combinations. These "behaviours" are difficult to elicit at this age and commonly captured through parent report. Whether one would expect the factors to predict vocabulary and word combinations depends on one's theoretical perspective. For some, the child's development from saying single words to combing two words together is a distinct stage marking the start of a grammatical system rather than a linear continuation in development. Although expressive language is clearly the most "manifest" element of early communication, considerable attention is also paid to verbal comprehension or receptive language. Partly, because it is commonly associated with general developmental levels but also because it is thought to be more important because it is predictive of subsequent communication (Chapman et al., 2000; Flax et al., 2009; Law et al., 2009). Parent report of comprehension is often considered to be less satisfactory than direct assessment because it can be very difficult for parents to judge how well their children understand in a highly contextualised communication environment (Law & Roy 2008; Styles & Plunkett 2009).

Two recent analyses have highlighted the relative role of social and parental characteristics in predicting language characteristics in very young children Hoff has suggested, using data comparing a relatively small sample of middle and higher SES groups, that differences in interaction patterns have a specific effect on vocabulary development and that more distal measures of SES add little to their models. Interestingly these differences did not feed through to comprehension or mean length of utterance (Hoff 2006). Since then, Morgan *et al.* (2015) have examined factors associated with two year old vocabulary in over six thousand children in the Early Childhood Longitudinal Program-Birth Cohort (ECLS B) study in the US and the potential implications that these factors have on reading, mathematics and behaviour at school entry. The authors concluded not only that two year vocabulary was a good indicator of later performance but that this was very sensitive to parental education and household income. They concluded that young children from higher socioeconomic status households experienced higher quality parenting at two years and had larger expressive vocabularies at 2 years compared with children from lower socioeconomic households.. It is important to add, as the authors acknowledge, that both parenting behaviours and vocabulary were recorded at the same time and the authors confined their focus to parent report of expressive vocabulary only and did not actually test the children's early receptive language skills or indeed the child's early grammatical skills or word combinations which have been suggested as a key indicator of risk for subsequent difficulties (Conti-Ramsden and Durkin 2016). Although the evidence suggests that parental behaviours are related to children's expressive language at 24 months, little research has addressed the degree to which parental behaviours are associated with other aspects of children's language development. Another concern is the timing of the point at which the prediction is made. It is not clear whether the

proximity of an experience to an outcome would be predicted to have a stronger association with that outcome than experiences recorded at an earlier stage.

One of the central issues is what should be done with information about preceding language skills and parenting behaviours from a practical point of view. A recent analysis has suggested that there may be a good case for integrating observation and assessment with social risk factors (McKean *et al.* 2016). In this study the identification of children based on a parent administered screening procedure focusing on early social interaction skills at 10-12 months was compared with a standard set of risk factors in terms of their ability to predict performance at four years. One of the key findings associated with a screening measure of social communication at 10-12 months was that the social gradient, so often reported, was not present at this early age. The results suggested that this early screening measure was able to identify groups of children who differed markedly in their language performance at four years but that this procedure was less successful at identifying which children would be "cases" in the sense of having language scores falling below a given threshold. But this pattern was improved once a series of risk factors (measured in terms of parental and child behaviours) were included in the models. It is noteworthy that in he 12 month data in this study both the gender balance (favouring girls) and the characteristic social gradient were not clearly demonstrated. Considerable interest in the role of language development has been expressed within a public health context (Law *et al.* 2013; Law & Pagnamenta 2017) but the evidence supporting universal interventions is not as strong as that for targeted interventions (Law *et al.*2017 but see also McGillion *et al.*2017; Blesses *et al.* .2018).

The aim of the study is to determine the level of association between expressive and receptive language at 15 months and language development at 2 years accounting for the contribution of proximal and distal measures of social disadvantage using the Avon Longitudinal Study of Parents and Children (ALSPAC). This is a large population-based birth cohort broadly representative of English children born in the early 1990s. It combines the strengths of both the Hoff and the Morgan et al., studies above by predicting to 24 months but including both receptive vocabulary (number of words understood) and expressive vocabulary (number of words produced) and expressive grammar (word combinations). It has the additional advantage of a directly administered standardised assessment of receptive language and a parent report measure of the child's ability to use word combinations as a key outcome at 24 months.

The study asks the following research questions:

- To what extent do proximal environmental factors predict language development at two years, adjusting for distal social factors?
- Does the pattern of association differ for receptive and expressive language (measured as vocabulary and early grammar)?

METHOD

Participants

ALSPAC is a prospective community ascertained population-based cohort study of all children born to mothers in an area of the west of England in the early 1990s, designed to explore the environmental and genetic factors that affect health and development. All mothers registering their pregnancy within the geographical county of Avon during the period from 1991-1992 were invited to participate. The eligible sample consisted of 20,248

pregnancies and the mothers of 14,541 (71.8% pregnancies were recruited ante-natally). Of these 14,541 pregnancies, 14,062 resulted in live births of which 13,988 were alive at one year of age (see Boyd *et al.* 2013 for a detailed description). The sample was found to have some under-representation of less affluent families and fewer families from black and ethnic minority groups than is the case nationally, although the overall developmental trajectories of the children were similar to national norms for the period (Roulstone *et al.* 2011). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee.

The Children in Focus (CiF) sample is a randomly selected sub-sample of the complete ALSPAC cohort. The purpose of the CiF sample was to collect direct observational and assessment data from the participants to both validate data collected via the parental questionnaires and reports collected in the full cohort, and also to collect data of several important developmental abilities across cognition, speech and language, physical development and motor ability. The CiF sample was chosen at random from the last 6 months of ALSPAC births occurring in 1992. The following exclusion criteria were adhered to: 1) mothers who had moved away from Avon or were no longer contactable; 2) no consent to participate; 3) infant death and; 4) very preterm infants (born less than 33 weeks). In addition to the parent and other report measures completed by the full ALSPAC cohort, the children in the CiF sample were invited to attend for clinic examination at 4, 8 and 12 months, and then at 6 monthly intervals up to the age of 61 months. In all, mothers of 2066 children from the full ALSPAC cohort were invited to participate in the clinics as part of the CiF sample. Of these, 1432 children (69%) including 18 sets of twins attended at least one clinic. At each time point, between 994 and 1314 children attended the examination clinics.

The CiF sample is the focus of the present analyses because all children in CiF completed a direct, face-to-face assessment of language comprehension at 2 years as well as parent report measures of vocabulary and grammar (word combinations). All assessments (direct and parent report) were carried out by fully qualified and trained staff (Roulstone *et al.* 2002).

Variables

All the variables are summarised and described in Table 1.

Biological risk: Two biological risk variables were selected as they are known to have an association with child development, including language development. These were gender, defined as male or female, and birthweight defined as the weight of the infant at birth and used as a proxy measure of developmental health. Both variables were obtained from parental questionnaire data collected at birth and in the first year of life.

Social disadvantage - proximal measures: A number of variables were selected reflecting what the parent did to stimulate their child, which the literature confirms to be associated with the child's language learning environment. A series of steps (including identifying variables of potential relevance, checking the original data collection measure and final content of the variables and eliminating non-relevant variables) were taken to identify variables that could be considered as influential in a child's developing language. A first conceptualisation of the possible components postulated three groups of variables, relating to: 1) the stimulation afforded to the child in the home; 2) the activities and opportunities offered to the child and the levels of environmental noise and; 3) the attitudes and emotional state of the mother and the perceived support available to the mother (Roulstone *et al.* 2011). The

items from these three groupings of variables were re-checked for completion and accuracy of codings, e.g., codings of variables were reversed where necessary to ensure the sense of direction across the variables were the same.

Measures in these groupings were derived from questionnaires mothers completed in the first 21 months of their child's life and are detailed below,

Group 1: The stimulation afforded to the child in the home.

Mothers completed questions about whether they try to teach their child, if they consider that babies need stimulation and a mother's parenting score, which was a derived variable asking mothers about the frequency with which they play with, sing to, share books with and engage their child in activities. Group 1 includes the HOME inventory (Bradley and Caldwell 1984). The Home Observation for Measurement of the Environment (HOME) is a widely used measure of the quality of the child's parenting and the home environment. The HOME inventory includes 21 items measuring a) parental activities including reading to the child, telling stories, singing and taking the child on errands to public places; b) having toys, records, books and audiotapes available in the home; and c) having a safe and supportive home environment.

Group 2: The activities and opportunities offered to the child and the levels of environmental noise.

Mothers completed questions about the number of books the child owns and if the child goes to the library, the number and type of toys the child owns, the frequency the TV is on in the home and if the child is looked after in a crèche.

Group 3: The attitudes and emotional state of the mother and the perceived support available to the mother.

Mothers completed the Edinburgh Postnatal Depression Questionnaire (Cox, Holden and Sagovsky 1987) and completed questions about the amount of support they receive in their role as a mother and their enjoyment of parenting.

The specific time points for these reported observations were between 6 and 21 months but these differed for the specific variables thus mother's parenting score and the HOME score

were rated at 6 and 18 months in group 1; the number of books owned was recorded at 6 and 18 months in group 2 and social support was rated at 8 and 21 months in group 3. The specific ages were dictated by the dataset but the key issue was that we wanted to assess these skills at more than one time point to ascertain whether there was an effect of timing. It is also important to acknowledge that these variables, while important in their own right, may also be closely associated to one another and may act as proxies for other behaviours. So, for example, recording that a child owns books does not necessarily mean that someone is reading those books with the child and TV being on in the home does not say anything about whether, or indeed how, the TV is being watched.

Distal Measures

A generic variable of social risk was constructed, following Schoon *et al.* (2004) as the distal measure of social disadvantage at birth and has been used elsewhere in our analyses (Roulstone et al., 2011; Clegg et al., 2015; Law et al., 2015. This consists of six measures; parental occupation (unskilled or manual: 0; skilled: 1), maternal education ('O' level or below (including vocational: 0; Higher than 'O' level: 1); housing tenure (Rented or other housing: 0; Owner occupied: 1); overcrowding (More than one person per room: 1; Less than one person per room: 0); financial difficulties (Financial difficulties: 1; None or minimal: 0) and use of a car (No: 0; Yes: 1). Each one was recoded into a binary variable as indicated, and then summed to provide a continuous 'social disadvantage' score for each child, with a range of 0-6 where 0 is a relatively disadvantaged social background and 6 represents a more advantaged background. Where children had either 1 or 2 items missing, scores were calculated proportionally from the other variables available.

Language

The final group of variables comprised a parent report of expressive and receptive vocabulary at 15 months. All mothers in the ALSPAC cohort were asked a series of questions about their child's understanding and use of language. The questionnaire was based on and modified from the MacArthur Communicative Development Inventory (CDI) (Fenson *et al.* 1993) Broadly the same questionnaire (same number of questions) was used at both the 15 month and 24 month time points. At the 15 month time point, the number of items was 134 and 123 at the 24 month time point (see table 1). This was due to modifications made to the questionnaire by the research team at the time. It is the same number of questions that was asked at the 15 month and 24 month time point. There are no differences in the number of questions or overlaps across the questions. The CDI has generally been shown to have good validity at a population level (Dale *et al.*1989; Feldman *et al.* 2006; Fenson *et al.* 2000) although concerns have been expressed at an individual level and especially with younger children (Feldman *et al*.2000).

Outcomes

There are three outcomes in this study. These were:-

i. Direct assessment of receptive language using the Reynell Developmental Language Scales (RDLS) Comprehension Scale at 25 months.

ii. Parent report of expressive vocabulary scale (adapted from the CDI see above)

iii. Parent report of expressive grammar – whether the child was reported to be using word combination at 24 months (adapted from the CDI see above)

The RDLS (Reynell 1977) is a standardised assessment used to assess language comprehension and expressive language. The RDLS comprehension scale measures a child's

language comprehension where the child is asked to respond to and carry out a series of spoken tasks using a standard set of toys (such as animals and figures as well as pictures). The assessment was administered and scored according to the assessment manual. The RDLS provided a direct assessment measure of language comprehension at 25 months for children in the CiF sample. The study refers to this assessment at 25 months as a 2-year outcome for the remainder of the paper.

Analysis

To address the two research questions, the study takes a standard modelling approach. In the univariable analysis stage, criteria is set to retain and exclude variables from this stage for the multivariable stage based on the magnitude and p value used (with the p value set at 0.1). This p value was considered a more liberal threshold so potentially important variables would not be excluded too early in the analysis. In the reporting of the results, care is taken not to over interpret the difference between statistically significant and non-significant findings.

Univariable regression analyses investigated the strength of the relationship with the 2-year language outcomes, i.e., the RDLS standardised score, the expressive vocabulary score and the expressive grammar (word combinations) score with a p-value of <0.1 (for reasons explained above). These are available in Appendix 1. Following this, a series of multivariable linear regression analyses investigated the relationships between the 2-year language outcomes and blocks of conceptually similar variables. In these multivariable analyses, where ordinary least squares regression was used for the comprehension and expressive vocabulary scores, and ordinal regression for the expressive grammar (word combinations) outcome, the more conventional threshold of $p<0.05$ was used to eliminate variables. Checks were made

for multi-collinearity and all were within accepted limits (Tolerance<0.2 Menard 1995; Variance Inflation Factor>1 Bowerman and O'Connell 1990). The sample for each analysis can vary as a result of missing data. Therefore, the sample size is shown for each regression analysis, which was conducted on an available data basis. To optimise the data, any information that was missing for the ALSPAC cohort members in the CiF sample was imputed. The method of imputation employed was multiple imputation by chained equations (MICE) implemented in STATA (Royston 2005). Five replicates of the data were created. Model estimates were averaged across these five analyses, with their standard errors calculated according to Rubin's rule (Rubin 1987). . Only small differences were found between the observed and imputed dataset but these were well within the limits of the confidence intervals and accordingly only the results from the observed dataset are reported here.

RESULTS

The descriptives for the full set of data included in the study are provided in Table 1. The mean number of words used were 61.42 (SD 34.52) whereas the number of words understood was considerably lower at 27.19 (SD 25.03) although as the standard deviations indicate the range was wide in both cases. Small positive correlations existed between social risk and the 2 year outcomes of comprehension($r=0.262$, $N=1050$, $p<0.001$), expression ($r=0.092$, $N=999$, $p=0.004$) and expressive grammar (word combinations), ($rho=0.107$, $N=978$, $p=0.001$) respectively). Thus the pattern of variability identified by Hart and Risley and many others is clearly present in these data but the differences are less pronounced than might be expected.

Table 1 about here

Results of the multivariable regression analyses

In these multivariable analyses, where ordinary least squares regression was used for the receptive and expressive vocabulary scores, and ordinal regression for the expressive grammar outcome (word combinations), the more conventional threshold of $p<0.05$ was used to eliminate variables. A series of models is generated for each of the outcomes; receptive language at 25 months, expressive vocabulary at 24 months and expressive grammar (word combinations) at 24 months.

Receptive language at 25 months

The results for this outcome are given in Table 2, which summarises a series of seven models with model seven as the final model. In the final model (model 7) after controlling for distal social risk, there was a small positive correlation for more words understood at 15 months, better parenting score at six months and the increasing 18 month HOME score on receptive language at 25 months. Similarly for gender difference, with girls performing slightly better than boys and TV watching, some days over every day, remained in this model. Books in the home, did not contribute to the comprehension outcome at 25 months. This final model accounted for 21.9% of the variance in the 2-year RDLS comprehension outcome.

Table 2 about here

Expressive vocabulary score at 24 months

The results for this outcome are given in Table 3, which summarises a series of eight models with model eight as the final model. After controlling for social risk (model 8), there is a profound gender effect with girls scoring ten words higher than the boys, and early language behaviours all remain in the final model. The latter are small and positive and probably

unsurprising given that the measure used is the same as that used at 24 months. Thereafter the 18 month HOME score accounts for all the remaining variance, with every unit increase implying an increase in expressive vocabulary of almost three. The final model accounts for 35.7% of the variance in the 2 year expressive vocabulary score.

Table 3 about here

Expressive grammar score (word combinations) at 24 months

Table 4 gives the results for the final outcome (word combinations) where 8 models are generated with model 8 as the final model. Again controlling for social risk, we find that there is an association for gender, in favour of girls with twice the odds of boys, and a small positive effect for better early language skills and HOME score at 18 months. Of those early language skills the effect is weak for both, explaining the lack of collinearity observed. Interestingly it is slightly stronger for early vocabulary (at fifteen months). The final model accounts for 27.4% of the variance in the expressive grammar score.

Table 4 about here

DISCUSSION

The current study examined the contribution of proximal measures of social disadvantage, adjusting for distal factors to children's early expressive and receptive language at 2 years while accounting for distal aspects of the environment and children's earlier language development. The study draws on data from over a thousand children in the ALSPAC CiF sample. The single strongest predictor was the relationship between gender and all the

outcomes, but especially expressive vocabulary. In general, predictors of verbal comprehension were less strong than those predicting expression perhaps reflecting the different method of assessment – direct assessment vs parent report. This emphasises the relative importance of social factors to expressive language performance. Regression models identified the contribution of distal (e.g., family income and material wealth) and proximal (what parents do with their children) measures of social disadvantage as well as early language development at 15 months to expressive vocabulary, expressive grammar (word combinations) and language comprehension development at 2 years of age. The overall variance explained for all three language outcome models was considerably less than 50%. More variance was explained in the 2-year expressive vocabulary outcome (35.7%) than the comprehension outcome (21.9%) and the figure for the ordinal regression for expressive grammar (word combinations) fell between the two (27.4%). Much of the variance (64.3%, 79.1% and 72.6% in each model respectively) is not accounted for and this suggests that other developmental or genetic mechanisms are likely to be contributing to language outcomes at 2 years. Nevertheless the first conclusion is that the models account for more variance than those described by Reilly and colleagues (Reilly et al.2007). Although there is no a priori reason for this difference it should make us cautious about generalising from single studies. One of the strengths of the present study is that it allows us to explore the potentially different mechanisms for both expressive and receptive outcomes..

The results suggest that the patterns of prediction from the different models vary considerably depending on the outcome. It is clear that our measure of social risk plays an important role for all three language outcomes. Nevertheless the environmental factors remain a factor in all the models albeit in different combinations (Hoff 2012). Our co-efficients are, as might be

anticipated, higher than in Hoff's smaller selected sample and rather lower than in the population studies (Morgan 2015). Thus, it is reasonable to argue that it is always the case that proximal factors add value to the models and to conclude that what parents/caregivers do with their child is important to children's language at 2 years of age. That said, it is also clear that the strong composite measures (the HOME inventory) accounts for much of the variance in other individual predictors leading them to drop out of our models, even if they are associated at a univarable level or at earlier stages of the models.

As earlier research would suggest, what might be termed pre-literacy activities in terms of number of books owned and library visits were important as was TV watching, specifically for the verbal comprehension. TV watching is a complex behaviour. Earlier research indicates that children under the age of 2 years are unlikely to benefit from children's TV (Close 2004), although a more recent re-analysis of the available data suggests that this may only be relevant at very high intensities (Foster and Watkins 2010). The TV question may be more to do with how parents/caregivers watch TV with their young children rather than the number of hours spent in front of a screen, a position made more complex nowadays by the almost universal access to other forms of media (laptops, tablets, games consoles and so on) at any time in most developed countries.

How a mother behaves with her child and the activities she undertakes are likely to be linked to her state of mind (Paulson *et al.* 2006). In the present analyses variables that reflected the mothers' enjoyment of parenting and psychological well-being dropped out early in the analyses. Rather it was often what the parents actually did that made the difference and, while it is assumed that maternal enjoyment of parenting and psychological well-being would

influence what parents/caregivers do with their children, this was not a finding from the study. Of particular interest is whether a mother's knowledge and beliefs of child development influences her behaviour with the child and the subsequent impact of this on child development. Much more needs to be more known about this dynamic interaction between maternal mental state and child development (Paulson *et al.* 2009).

We were interested in the timing of the variables. Unsurprisingly early language variables at 15 months also made an important contribution to the 2-year language outcomes specifically the 2-year expressive vocabulary outcome and 2-year expressive grammar (word combinations) outcome but less so for the 2-year comprehension outcome. Early language development in the first 15 months is therefore important in determining language development at 2 years and raises some interesting questions about if and how language development at this age could be screened to identify children potentially at risk for language delay and subsequent more pervasive speech, language and communication impairments. There is, of course, a technical issue about when such behaviours are reported and observed and how strong the measures are given the nature of the assessment of the skills of very young children and given the likelihood that the closer the measures are to one another the greater the likelihood that they will be associated. The more valid and reliable a measure is, the greater the likelihood that it will measure the same construct at a later date – especially if that date is not very far from the point that the predictor was measured. It is noteworthy that the parent report of language measure remains in all the models and is consistently stronger for parent reported outcomes although it is important to note that parent report of comprehension appears to have a specific association (albeit lower) with tested receptive language performance. The other caveat in this context is that the parent report measures of

language in the ALSPAC sample are not identical, the first being longer and, as the average scores in Table 1 indicate, rather it may easier to over report a child's understanding at the earlier stages of development but this become increasingly difficulty with time. Indeed the measurement challenges in the younger age groups has been identified elsewhere (Feldman *et al.* 2000), This will not affect the direction of the association but it may affect its size. There is a suggestion that even earlier parent report of infant communication (ie at 12 months) predicts later tested language skills (McKean *et al* 2016) and one could argue that the wider the gap in age between predictor and outcome the more convincing the argument for the association would be. The two other variables that allow us some insight into timing are book ownership and the mother's parenting score both recorded at six and eighteen months. The earlier parenting score is more highly associated with comprehension whereas the later parenting score is significant on its own but then drops out of the model. Parenting at both time points predict expressive vocabulary (the earlier stronger than the later) and only the later time point predicts expressive grammar but then they drop out of the models altogether. Similarly book ownership, at both ages, predicts all aspects of language as they enter the model but with a stronger association for expressive grammar and vocabulary but in this case the later time point is always a stronger predictor. However, this pattern does not persist in the final modules as these variables drop out in the context of other proximal variables.

The study by Morgan et al (2015) is the most comparable to this present study. Although, the construction of the models in the two studies differ, a similar pattern of findings is identified. In the Morgan et al., study, the role of SES was weaker but , the parenting score prediction was stronger than it was in our data but our parenting measures were taken sometime before the language measures were taken which is likely to affect the predictive power of the

variable. Although it is technically possible there is a prediction between two variables taken at the same time point we would argue it is the prospective rather than the predictive significance which is key here. By contrast, the TV watching variable was only significant for comprehension at the univariable level and a multivariable level (0.3) but not significant for vocabulary or expressive grammar (word combinations) (0.19 0.42,0.34) comparable to that in Morgan's (-0.01/-0.4).

Practical implications

These findings give a clear indication that combining earlier language measures with proximal risk factors has the potential to predict early communication behaviours with potential implications for the early identification of children at risk of slower language development. In the Morgan et al. study the final co-efficient for general cognitive functioning was 0.61, much higher than all the other variables put together. But realistically the use of a comprehensive standardised measure of child development is not feasible in a publication health context. By contrast, our vocabulary checklist at 15 months is feasible for public health practitioners to use. These findings suggest that improving parent/child learning opportunities is likely to contribute to two year outcomes and this, in turn, is likely to lead to longer term effects. In some cases these appear to be fairly specific, for example, the relationship between parenting in general and book reading in particular at 8 months and expressive grammar (word combinations) at 24 months. In most other situations the amount of variance accounted for by specific measures is relatively small and they need to be taken together to have an effect. There are clearly some important messages here for public health professionals and there is evidence that speech and language therapists are working closely with health visitors and in the shaping such messages (Law and Pagnamenta 2017). Whether these results could be translated into a screening measure with functional validity is another matter. To date such behaviour/risk combined measures have shown merit but have proved insufficiently accurate to predict case status (of language difficulties) over time (McKean et al. 2016).

Methodological limitations

ALSPAC is a highly respected, community ascertained, cohort study (Sonuga-Barke 2011). Focusing on the CiF sample rather than the whole ALSPAC cohort enabled the inclusion of the direct standardized language assessment of comprehension at 2 years along with the parent report measures. The CiF sample consisted of between 994 and 1314 children, and in the present study the number of participants in the regression analyses ranged from 892 and 1087. Indeed, the sample size in the comprehension outcome regression analysis was marginally smaller than in the expressive vocabulary and expressive grammar analyses. It is recognised that the size of the CiF sample used may not be large enough to be representative of the ALSPAC cohort although random selection of participants and other measures to increase representativeness were addressed.

As expected, the ALSPAC sample suffered from attrition, which is common across longitudinal cohort studies. Nevertheless there is always a concern that the more socially disadvantaged sections of the population are the ones most likely to drop out and indeed this is the case for the CiF sample. The fact that the multiple imputation analyses replicated the findings from the multivariable regression analyses suggests that the effects of potential bias and attrition are relatively limited. Nonetheless it is conceivable that the variance might have increased and the effects identified in the models might be stronger if these groups had been included. The limitations of the measures used are acknowledged in terms of parent/caregiver report of children's language abilities as well as the proximal measures of social disadvantage and potential overlaps between these variables. This raises important questions about the potential trade-off between cost efficiency and robustness in capturing the language ability of very young children in general population cohort studies. Furthermore, general cohort

population studies are subject to changing societal influences. For example, compared with the 1990s, library provision in the UK has been considerably reduced, and the ways children and parents/caregivers access and read books, and watch TV, has changed markedly – and so this study does not necessarily reflect current social and cultural practices in these domains. Finally, it is important to acknowledge that the data presented in this paper are associations rather than causal in nature. Trials of interventions targeting key risk behaviours would be needed to determined whether these are causal mechanisms and, although there are a number of effective interventions targeting early child development, these rarely map directly on to the observational data.

CONCLUSIONS

These findings confirm the importance of the association between the child's earliest experiences and their early language development but they go further than Morgan et al. (2015) in allowing us to tease apart some of the specific mechanisms for our three different outcomes. Our social risk composite remained in all the models but other factors associated with the child's environment continued to operate above and beyond structural social factors. Predictions are strongest for parent report of expressive vocabulary and expressive grammar (word combinations) and weakest for comprehension. Similarly social factors both (distal and proximal) are much stronger for vocabulary and for word combinations than they are for comprehension. It is important to note that the parenting score and TV watching appear to be especially important in their association with comprehension but not the other skills. The fact that earlier vocabulary remains in all the models suggest this as a potentially modifiable mechanism which is of direct relevance to those working with very young children. Indeed, our results suggest that the predictive role of language becomes clear by as early as 15

months, indicating that this is a critical time to identify those most at risk. Parents/caregivers in contexts of social disadvantage should be encouraged and supported to facilitate their children's language development from birth but there is a pressing need to develop the evidence base underpinning early interventions which target these skills.

REFERENCES

BLESES, D. , HøJEN, A. , JUSTICE, L. M., DALE, P. S., DYBDAL, L. , PIASTA, S. B., MARKUSSEN-BROWN, J. , CLAUSEN, M. and HAGUISH, E. F. (2018), The Effectiveness of a Large-Scale Language and Preliteracy Intervention: The SPELL Randomized Controlled Trial in Denmark. Child Dev, 89: e342-e363. doi:10.1111/cdev.12859

BOWERMAN, B.L and O'CONNELL, R.T. 1990. Linear statistical models: an applied approach (2nd ed.) Boston: PWS-Kent Publishing Company.

BOYD, A., GOLDING, J., MACLEOD, J., LAWLOR, D.A., FRASER, A., HENDERSON, J., MOLLOY, L., NESS, A., RING, S. and DAVEY SMITH, G. 2013, Cohort profile: the 'children of the 90s' – the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, 42, 111-127.

BRADLEY, R. H., 1993, Children's home environments, health, behavior, and intervention efforts: A review using the HOME inventory as a marker measure. *Genetic, Social and General Psychology Monographs*, 119, 437–490.

BRADLEY, R.H., and CAULDWELL, B. 1984, Home Observation for Measurement of the Environment (HOME) – revised edition, Little Rock, Ar, University of Arkansas, Little Rock, USA.

CHAPMAN, R.S., SEUNG, H., SCHWARTZ, S.E. and BIRD, E. 2000. Predicting language production in children and adolescents with Down Syndrome. *Journal of Speech, Language 7 Hearing Research,* 43, 340-350.

CLEGG, J., LAW, J., RUSH, R., PETERS, T. and ROULSTONE, S. 2015. The contribution of early language development to children's emotional and behavioural functioning at 6 years: an analysis of data from the children in focus sample from the ALSPAC birth cohort. *Journal of Child Psychology & Psychiatry*, 56, 1, 67-75.

CLINTON FOUNDATION (2013) preparing America's children for Success in the 21st century: Too Small to fail. New York: Clinton Foundation https://www.clintonfoundation.org/blog/2013/10/03/closing-word-gap Accessed 1.17)

CLOSE, R., 2004, *Television and language Development in the Early Years: a review of the literature*. London: National Literacy Trust.

CONTI-RAMSDEN, G. 7 DURKIN, K. 2015, What factors influence language impairment: Considering resilience as well as risk *Folia Phoniatr Logop* 67:293-299 doi.org/10.1159/000444750.

COX, J.L., HOLDEN, J.M., and SAVOGSKY, R., 1987, Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry* 150,782-786.

DALE, P., BATES, E., REZNICK, J., and MORISSET, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language, 16*(2), 239-249.

FLAX, J.F., REALPE-BONILLA, T., ROESLER, C., 2008, Using early standardised language measures to predict later language and early reading outcomes. *Journal of Learning Disabilities,* 42, 1.

FELDMAN, H. M., DOLLAGHAN, C. A., CAMPBELL, T. F., KURS-LASKY, M. , JANOSKY, J. E. and PARADISE, J. L. 2000, Measurement Properties of the MacArthur Communicative Development Inventories at Ages One and Two Years. Child Development, 71: 310-322. doi:10.1111/1467-8624.00146.

FELDMAN, H. M., DALE, P. S., CAMPBELL, T. F., COLBORN, D. K., KURS-LASKY, M., ROCKETTE, H. E. and PARADISE, J. L. 2005, Concurrent and Predictive Validity of Parent Reports of Child Language at Ages 2 and 3 Years. *Child Development,* 76: 856–868. doi:10.1111/j.1467-8624.2005.00882.x.

FENSON, L., DALE, P.S., REZNICK, J.S., THAL, D., Bates, E., HARTUNG, J.P., PETHICK, S. and REILLY, J.S., 1993, *The MacArthur Communicative Development Inventories: user's guide and technical manual*. Singular Publishing Group, San Diego.

FENSON L., BATES, E., DALE, P., GOODMAN, J., REZNICK, J. S. and THAL, D. 2000, Reply: Measuring Variability in Early Child Language: Don't Shoot the Messenger. *Child Development*, 71: 323–328. doi:10.1111/1467-8624.00147.

FOSTER, E.M. and WATKINS, S.,2010. The value of re-analysis; TV viewing and attention. *Child Development* 81, 368-375.

HART B. and RISLEY, T., 1995, *Meaningful Differences in the Everyday Experiences of young American Children*. Baltimore: Paul Brookes.

HOFF, E., 2006, How social contexts support and shape language development. *Developmental Review*, 26, 55-88.

HOFF, E. 2012 Causes and consequences of SES-related Differences in parent-to-child Speech in M.Bornstein and Robert Bradley Eds Socio-economic status parenting and child development. (pp. 147-160). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

HUTTENLOCHER, J., WATERFALL, H., VASILYEVA, M., VEVEA, J., HEDGES, L.V., 2010, Sources of variability in children's language growth. *Cognitive Psychology*, 61, 343-365.

JESSOR, R. and JESSOR, S.L. (1973) The Perceived Environment in Behavioral Science Some Conceptual Issues and Some Illustrative Data *American Behavioral Scientist* 16): 801-828 https://doi.org/10.1177/000276427301600602.

KARRAS, J. and BRAUNGART-RIEKAR, J.M., 2005, Effects of shared parent-infant book reading on early language acquisition. *Applied Developmental Psychology*, 26, 133-148.

KRIEGER,N. 2008,Proximal, Distal, and the Politics of Causation: What's Level Got to Do With It? *American Journal of Public Health* 98, 221-230.

LAW, J. CHARLTON, J., DOCKRELL, J., GASCOIGNE, M., McKEAN, C. and THEAKSTON, A. (2017) Early Language Development: Needs, provision and intervention for preschool children from socio-economically disadvantage backgrounds. London: Education Endowment Foundation.

LAW, J. & PAGNAMENTA, E. (2017) Public Health Interventions: Promoting the development of young children's language *Bulletin of the Royal College of Speech and Language Therapists* 777 January12-15.

LAW, J., RUSH, R., CLEGG, J., PETERS, T., and ROULSTONE, S., (2015), The role of pragmatics in mediating the relationship between social disadvantage and adolescent behaviour. *Journal of Developmental and Behavioural Paediatrics*, 36, 389-398.

LAW, J**.,** REILLY, S. and SNOW, P. (2013) Child speech, language and communication need in the context of public health: A new direction for the speech and language therapy profession. *International Journal of Language and Communication Disorders* DOI: 10.1111/1460-6984.12027.

LAW, J., MCBEAN, K. and RUSH, R., 2011, Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. *International Journal of Language and Communication Disorders*, 46, 657-664.

LAW, J., RUSH, R., SCHOON, I. and PARSON,S. (2009). Modeling developmental language difficulties from school entry age into adulthood: literacy, mental health and employment outcomes. *Journal of Speech, Language & Health Research*, 52, 1401-1416.

LAW, J., and ROY, P. 2008, Parental report of infant language skills: a review of the development and application of the communication development inventories. *Child & Adolescent Mental Health*, 13, 4, 198-206.

McGILLION, M. , PINE, J. M., HERBERT, J. S. and MATTHEWS, D. (2017), A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *J Child Psychol Psychiatr*, 58: 1122-1131. doi:10.1111/jcpp.12725.

McKEAN, C., LAW, J., MENSAH, F., CINI, E., EADIE, P. and FRAZER, K. (2016) Predicting meaningful differences in school-entry language skills from child and family factors at one year *International Journal of Early Childhood* DOI 10.1007/s13158-016-0174-0.

MENARD, A., 1995, Applied logistic regression analysis. *Sage university paper series on quantitative applications in the social sciences, 07-106*. Thousand Oaks, CA: Sage.

MORGAN,P.L., FARKAS, G., HILLEMEIER, M.M., SCHEFFNER HAMMER, C. and MACZUGA, S., 2015, 24 month old children with larger oral vocabularies display greater academic and behavioural functioning at kindergarten entry. *Child Development*, 86, 1351-1370.

PAULSON, J.F., DAUBER, S. and LEIFERMAN, J.A., 2006, Individual and combined effects of postpartum depression in mothers and fathers on parenting behaviour. *Pediatrics*, 118, 659-668.

PAULSON, J.F., KEEFE, H.A. and LEIFERMAN, J.A., 2009, Early parental depression and child language development. *Journal of Child Psychology & Psychiatry*, 50, 254-262.

REILLY,S., WAKE, M., BAVIN,E.L., PRIOR, M., WILLIAMS, J., BRETHERTON, L., EADIE,P. BARRETT,Y., UKOUMUNNE, O.C. (2007) Predicting Language at 2 Years of Age: A Prospective Community Study Pediatrics 120 (6) e1441-e1449; DOI: 10.1542/peds.2007-0045.

RESCORLA, L., ROSS, G.S., and MCCLURE, S. Language delay and behavioural/emotional problems in toddlers: findings from two developmental clinical. *Journal of Speech, Language & Hearing Research*, 50, 1063-1978.

REYNELL, J., 1977, *The Reynell Developmental Language Scales (Revised)*. Windsor: NFER Publishing Company, UK.

ROY, P., and CHIAT, S., 2013, Language and socioeconomic disadvantage: teasing apart from delay and deprivation from disorder: In Marshall, C. (ed), Current issues in developmental disorders, Hove, Psychology Press, pp 125-150.

ROULSTONE, S., LAW, J., CLEGG, J., RUSH, R. and PETERS, T.J., 2011, *Investigating the role of language in children's early educational outcomes*. Department for Education. Research Report DFE-RR134.

ROULSTONE, S., LOADER, S., NORTHSTONE, K., BEVERIDGE, M., and THE ALSPAC TEAM, 2002, The speech and language of children aged 25 months: descriptive data from the Avon Longitudinal Study of Parents and Children. *Early Child Development and Care*, 72, 259-268.

ROYSTON, P., 2005, *Multiple imputation of missing values: update of ICE*. Stata Journal, 5, 527, 536.

RUBIN, D.D., 1987, *Multiple imputation for nonresponse in surveys*. New York, USA:Wiley.

SCHOON, I. (2006). *Risk and resilience: adaptations in changing times*. Cambridge: Cambridge University Press.

SCHOON, I., PARSONS, S. and SACKER, A, 2004, Socioeconomic adversity, educational resilience and subsequent levels of adult adaptation. *Journal of Adolescent Research*, 19, 383- 404.

SONUGA-BARKE, E.J.S., 2012, Editorial: the developmental psychopathologist as scientist-sleuth – can large-scale longitudinal birth cohort studies provide the missing clues? *Journal of Child Psychology and Psychiatry*, 53, 619-621.

SPENCER, S., CLEGG, J. and STACKHOUSE, J., 2012, Adolescents from differing socioeconomic backgrounds: a comparison of language skills. *International Journal of Language and Communication Disorders*, 47, 274-284.

STYLES, S. and PLUNKETT, K., 2009, What is word understanding for the parent of a one year old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, 36, 2, 895-908.

WHITEHOUSE, A.J., ROBINSON, M. and ZUBRICK, S.R., 2011, Late talking and the risk for psychosocial problems during childhood and adolescence. *Paediatrics*, 128, e324-332.

Table 1 Descriptive statistics for included variables (frequency (%) and mean scores (SD))

Variables		**Total sample**
Biological		**N (%)**
Gender (n=1087)	Male	593 (54.6)
	Female	494 (45.4)
Birthweight (kg) (n=1077)		3.46 (0.51)
Distal socio-economic status measure		**Mean (SD)**
Social risk (n=1050)		4.55(1.30)
Receptive and expressive language at 15 and 2 years		**Mean (SD)**
Number of words child understands (out of 134) at 15 months (n=1019)		74.76 (30.36)
Number of words child can say (out of 134) at 15 months (n=1019)		13.99 (17.11)
Number of words child understands (out of 123) at 24m (n=1023)		27.19 (25.03)
Number of words child can say (out of 123) at 24 months (n=1023)		61.42 (34.52)
RDLS comprehension score at 25 months (n=1087)		24.00 (8.43)
Expressive grammar at 24 months (n=1000)		**N (%)**
	Not yet	179 (17.9)
	Sometimes	299 (29.9)
	Often	522(52.2)
Proximal socio-economic status measure		
Group 1 Stimulation at home		**N (%)**
Mother tries to teach child (n=1055)	Yes	813(77.1)
	No	242 (22.9)
“Babies need stimulation” (n=1042)	Agree	1030 (98.8)
	Disagree	12 (1.2)
		Mean (SD)
Mother's parenting score at 6 months (n=1051)		10.59 (1.41)
Mother's parenting score at 18 months (n=1048)		40.87 (4.44)
HOME score at 6 months (n=1055)		8.41(2.12)

HOME score at 18 months (n=1048)		10.49 (1.55)
Group 2 Activities, opportunities and environmental noise		
Child looked after in crèche at 24 months (n=1007)	No	883 (87.7)
	Yes	124 (12.3)
Number of books child owns at 6 months (n=1054)	2 or less	364 (34.5)
	>2	690 (65.5)
Number of books owned by child at 18 months (n=1048)	2 or less	40 (3.8)
	>2	1008 (96.2)
Child goes to library at 18 months (n=1040)	Once Per Month or less	964 (92.7)
	At least once Per Week	76 (7.3)
Frequency the TV is on reported at 18 months (n=1006)	Yes, Every Day/Some days	822 (81.7)
	No, Hardly Ever	184 (18.3)
Group 3: Attitudes and emotional state of the mother and support		**Mean (SD)**
Social support score at 8 months (n=993)		20.44 (4.86)
Social support score at 21 months (n=934)		21.15(4.86)
Edinburgh Postnatal Depression Questionnaire (n=1053)		24.23 (4.58)
Maternal enjoyment of parenting score (n=1032)		13.3 (2.12)

Table 2 Multivariable regressions with comprehension at 25 months as the outcome (regression coefficient and 95% CI)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	N=1087	N=1050	N=1019	N=1022	N=987	N=932	N=934
Biological							
Gender (male/female)	0.307*** (0.188, 0.425)					0.236*** (0.12, 0.352)	0.235*** (0.12, 0.35)
Distal socio-economic status measure							
Social Risk		0.200*** (0.156, 0.245)				0.150*** (0.104, 0.196)	0.159*** (0.114, 0.203)
Receptive and expressive language at 15 months							
Number of words child understands (out of 134) at 15 months			0.010*** (0.008, 0.012)			0.009*** (0.007, 0.011)	0.009*** (0.007, 0.011)
Proximal socio-economic status measures							

Group 1 Stimulation at home							
Mother's parenting score, 6 months				0.085*** (0.038, 0.132)		0.055* (0.008, 0.103)	0.062** (0.018, 0.106)
Mother's parenting score, 18 months				0.019* (0.004, 0.034)		0.003 (-0.013, 0.018)	
HOME score, 18 months				0.123*** (0.081, 0.166)		0.076** (0.031, 0.12)	0.090*** (0.05, 0.13)
Group 2 Activities, opportunities and environmental noise							
Number of books child owns at 6 months					0.296*** (0.166, 0.427)	0.059 (-0.073, 0.191)	
Number of books owned by child at 18 months					0.688*** (0.36, 1.016)	0.204 (-0.125, 0.533)	
Frequency the TV is on reported at 18 months					0.246** (0.09, 0.402)	0.186* (0.038, 0.334)	0.19* (0.042, 0.337)
R square	0.023	0.069	0.094	0.098	0.06	0.222	0.219
p<.05* p<.01** p<.001***							

Table 3 Multivariable regressions with expressive vocabulary at 24 months as the outcome (regression coefficient and 95% CI)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
	N=1023	N=999	N=990	N=983	N=989	N=954	N=892	N=958
	B 95% CI	B 95% CI	B 95% CI	B 95% CI	B 95% CI	B 95% CI	B 95% CI	B 95% CI
Biological								
Gender (male/female)	15.711*** (11.564, 19.859)						9.272*** (5.547, 12.996)	10.028*** (6.453, 13.604)
Distal socio-economic status measure								
Social Risk		2.426** (0.786, 4.067)					2.154** (0.619, 3.689)	2.838*** (1.465, 4.212)
Receptive and expressive language at 15 months								
Number of words child understands (out of 134) at 15 months			0.260*** (0.191, 0.329)				0.227*** (0.154, 0.3)	0.228*** (0.159, 0.296)
Number of words child can say (out of 134) at 15m			0.825*** (0.699, 0.950)				0.779*** (0.646, 0.913)	0.808*** (0.681, 0.935)
Proximal socio-economic status measure								
Group 1 Stimulation at home								
Mother's parenting score, 6 months				2.004* (0.329,			1.238 (-0.326,	

				3.679)			2.803)	
Mother's parenting score, 18 months				0.916** (0.369, 1.463)			0.075 (-0.435, 0.584)	
HOME score, 18 months				4.380*** (2.867, 5.893)			1.443 (-0.018, 2.904)	2.562*** (1.354, 3.769)
Group 2 Activities, opportunities and environmental noise								
Number of books child owns at 6 months					10.233*** (5.667, 14.799)		3.309 (-0.99, 7.607)	
Number of books owned by child at 18 months					18.001** (6.263, 29.738)		4.785 (-6.421, 15.99)	
Group 3: Attitudes and emotional state of the mother and support								
Social support score at 8 months						0.512* (0.06, 0.963)	0.099 (-0.296, 0.495)	
R square	0.051	0.008	0.303	0.098	0.035	0.005	0.348	0.357
p<.05* p<.01** p<.001***								

Table 4 Ordinal Regression with expressive grammar (word combinations) at 24 months as the outcome (OR 95% CI)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model
	OR	OR	OR	OR	OR	OR
	(95%CI)	(95%CI)	(95%CI)	(95%CI)	(95%CI)	(95%C
	N=1087	N=1087	N=1087	N=980	N=968	N=889
Biological						
Gender (male/female)	2.303*** (1.802, 2.945)					
Distal – Socio-economic status variable						
Social risk		1.172** (1.069, 1.284)				
Language variables at 15 months						
Number of words child can say (out of 134) at 15 months			1.082*** (1.067, 1.099)			
Proximal socio-economic variables						
Group 1: Stimulation at home						
Mother's parenting 18 months				1.042** (1.011, 1.075)		
HOME score 18 months				1.267*** (1.163, 1.381)		
Group 2: Activities, opportunities and environmental noise						
Number of books owned 6 months					1.320* (1.020, 1.709)	
Number of books owned at 18 months					2.450** (1.281, 4.688)	
Group 3: Attitudes and emotional state of the mother and support						
Social support score at 21 months						1.026* (1.001 1.053)
Model						
-2 Log Likelihood	26.117	83.326	326.25	604.043	40.955	199.27
Chi-Square (df)sig	45.768(1), p<.001	11.428(1), p=.001	196.299(2), p<0.001	60.567(2), p<0.001	15.525(2), p<0.001	4.088(1 p=.043
Nagelkerke	0.052	0.013	0.212	0.069	0.018	0.005

p<.05* p<.01** p<.001***